D388–D393 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 28 October 2011*
doi:10.1093/nar/gkr911

ProGlycProt: a repository of experimentally characterized prokaryotic glycoproteins

Aadil H. Bhat[1], Homchoru Mondal[1], Jagat S. Chauhan[2], Gajendra P. S. Raghava[2], Amrish Methi[3] and Alka Rao[1,*]

[1]Protein Science and Engineering, Institute of Microbial Technology, Council of Scientific and Industrial Research, India [2]Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research, India and [3]SS Compusoft Private Limited, Jaipur, Rajasthan, India

Received August 14, 2011; Revised September 30, 2011; Accepted October 8, 2011

ABSTRACT

ProGlycProt (http://www.proglycprot.org/) is an open access, manually curated, comprehensive repository of bacterial and archaeal glycoproteins with at least one experimentally validated glycosite (glycosylated residue). To facilitate maximum information at one point, the database is arranged under two sections: (i) ProCGP—the main data section consisting of 95 entries with experimentally characterized glycosites and (ii) ProUGP—a supplementary data section containing 245 entries with experimentally identified glycosylation but uncharacterized glycosites. Every entry in the database is fully cross-referenced and enriched with available published information about source organism, coding gene, protein, glycosites, glycosylation type, attached glycan, associated oligosaccharyl/glycosyl transferases (OSTs/GTs), supporting references, and applicable additional information. Interestingly, ProGlycProt contains as many as 174 entries for which information is unavailable or the characterized glycosites are unannotated in Swiss-Prot release 2011_07. The website supports a dedicated structure gallery of homology models and crystal structures of characterized glycoproteins in addition to two new tools developed in view of emerging information about prokaryotic sequons (conserved sequences of amino acids around glycosites) that are never or rarely seen in eukaryotic glycoproteins. ProGlycProt provides an extensive compilation of experimentally identified glycosites (334) and glycoproteins (340) of prokaryotes that could serve as an information resource for research and technology applications in glycobiology.

INTRODUCTION

Protein glycosylation in prokaryotes is a recent but rapidly growing area of research. An expanding repertoire of prokaryotic glycoproteins is increasingly being explored as a target for therapeutic interventions in diagnostics (1), vaccines (2), as future nano-machines using proteins like S layer glycoproteins (3) and as a strategy to improve industrially important enzymes for specific attributes (4,5).

The prokaryotes indeed synthesize a wide variety of glycans linked covalently to their proteins, commonly at the amide group of Asn (N-linked), hydroxyl group of Ser/Thr/Tyr (O-linked) and rarely at the sulphur residue of Cys (S-linked) (6). Equally, they display a diversity in the mechanisms of glycosylation that include well-known, *en bloc* N-glycan transfer (Archaea & *Campylobacter* spp.) and sequential O-glycan transfer (*Pseudomonas* spp., *Campylobacter* spp. etc.) as well as novel, *en bloc* O-glycan transfer (*Neisseria* spp.) and sequential N-glycan transfer (*Haemophilus influenzae*) (7,8). Accordingly, it has led to identification and characterization of several new protein glycosylation-associated enzymes, OSTs and GTs in prokaryotes (7,8). Likewise, hundreds of new glycoproteins have now been identified experimentally, across all major phyla of bacteria and archaea (Supplementary Figure S1), implicating them in diverse biological functions in cellular and extra cellular milieu (9). To name a few, Apa protein of human pathogen *Mycobacterium tuberculosis* (10), flaA of phytopathogen *Acidovorax avenae* K1 H8301 (11), glycosylated pilin protein of *Neisseria gonorrhoeae* (12), and adhesins of several pathogenic bacterial species are the examples of glycoproteins that are involved in crucial host–pathogen interactions, modulation of the host immune system and virulence of the pathogenic bacterial species. Interestingly, in the last decade, as many as 67 new glycoproteins have been characterized for their glycosites in prokaryotes (Figure 1). Around the same time, many reviews and

*To whom correspondence should be addressed. Tel: +91 172 6665214; Fax: +91 172 2690585; Email: raoalka@imtech.res.in

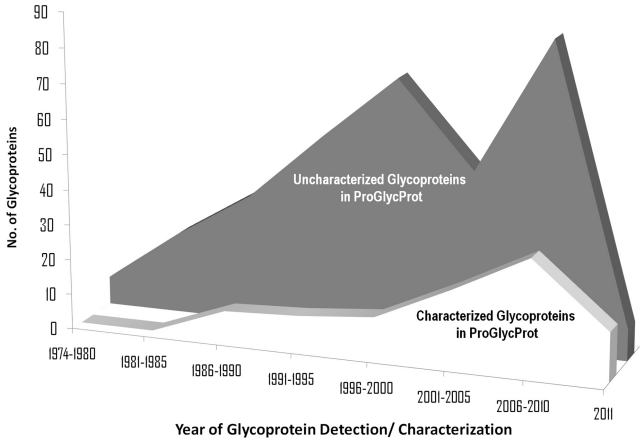

Figure 1. Trends of experimental research on prokaryotic glycoproteins in last 35 years as derived from ProGlycProt database.

research articles have appeared in reputed scientific journals containing focused compilations of known information about these glycoproteins (3,9,13–16, http://www.proglycprot.org/recent_review.aspx). The rise in the interest in glycoproteins and glycobiology of prokaryotes is obvious. However, currently, there is no specialized resource for prokaryotic glycoproteins providing information in a comprehensive manner. Also, a dedicated resource for prokaryotic glycoproteins analogous to O-GLYCBASE (a collection of O- and C-glycosylated proteins of eukaryotes, 17) will complement the ongoing efforts of glycoprotein annotation as at Swiss-Prot (18) and the one like dbPTM, integrating experimentally validated information on post-translational modifications (19). Further, with the availability of high-throughput techniques like mass spectroscopy, lectin arrays and emerging data analysis tools, a large influx of data on prokaryotic glycoproteins are anticipated.

In view of this necessity, and to cater to general interests in the science of prokaryotic glycoproteins, we have developed ProGlycProt as a manually curated, comprehensive repository of published information on bacterial and archaeal glycoproteins with at least one experimentally characterized glycosite. It is a modest but focused beginning of an effort to provide enough experimental information at one point, to glean insights into the relationship between a glycoprotein, its OSTs/GTs, protein glycosylation-linked gene (s) and their genomic context.

In this database, a characterized glycoprotein is the one where at least one glycosite is validated through experiments like Edman degradation, mass spectroscopy or site-directed mutagenesis. Similarly, an uncharacterized glycoprotein is the one, where glycosylation but not glycosite (s) is identified by one or more experimental methods, e.g. aberrant migration on SDS–PAGE, sugar specific staining, lectin binding, etc.

DATA COLLECTION AND CURATION

The first release of ProGlycProt with 340 entries is a result of an extensive literature search followed by the manual curation of the data compiled from a total of 410 research articles and review papers (http://www.proglycprot.org/Bibliography.aspx). For ProCGP, the initial literature collection was built using various keyword searches made at Pubmed (20), Google Scholar and the Web of science. Additional references relevant to this study were retrieved from the citations given in aforementioned research and review articles. As a result, ProCGP now lists 88 native glycoproteins, in addition to seven proteins and peptides that are glyco-engineered using *in vitro*/*in vivo* and enzymatic or synthetic approaches. ProCGP represents all three experimentally known protein-glycan linkages in prokaryotes, namely N, O and S with information on 132 N-glycosites, 196 O-glycosites and 6 S-glycosites (Supplementary Figure S2). Both identical (five proteins with 18 glycosites are identical in the current database) and homologous sequences are included to provide a complete primary list of experimentally characterized prokaryotic glycoproteins from which a non-redundant dataset can be derived easily as required by the users. In some cases, a redundant entry may provide interesting experimental information. For example, ProGlycProt ID AC102 provides information on *in vivo* N-glycosylation at noncanonical sequon **NX(N/L/V)** (X≠ P) in engineered mutants of a cell surface glycoprotein (CSG/S layer glycoprotein derived from AC101) at position N36 in the full-length protein by a yet unknown OST in archaea *Halobacterium salinarum* [known as *H. halobium* previously (16,21)]. Similarly, identical entries BC130, 132, 133, 135 and 136 are included as each belongs to a different strain. First, all entries in ProCGP are manually corrected for incorporation of mutational changes/sequence conflicts/engineered sequences, if any, as per the experimental data and later annotated for experimentally verified glycosites. A visual display of these manually annotated sequences is available under subfield titled 'glycosite (s) annotated protein sequence'. Therefore, this field is a true identifier for redundancy estimation in the database. The glycoprotein entries (21 in number) retrieved initially from Swiss-Prot to nucleate data-section ProCGP are revised as per the updated literature in applicable cases like S-layer glycoprotein of *H. salinarum,* S-layer protein of *Haloferax volcanii* and AIDA auto-transporter protein of *Escherichia coli*. A sequence conflict is addressed for HisJ protein of *Campylobacter jejuni*. Finally, a cross-check with BCSDB version 3.0 (22), O-GLYCBASE version 6.0 and Swiss-Prot release 2011_07 suggests that ProCGP is a comprehensive, exclusive and currently the largest compilation of characterized prokaryotic glycoproteins and their glycosites (Supplementary Tables S1–S3).

In parallel, cataloguing of uncharacterized prokaryotic glycoprotein entries was made under data-section ProUGP from independent reviews and research articles published in various journals as mentioned in the Introduction. Nonetheless, ProUGP contains at least 107 experimentally identified glycoprotein entries from prokaryotes (with unsequenced genomes) that are not available in Swiss-Prot release 2011_07.

DATA ARRANGEMENT AND ACCESS

ProGlycProt is developed by integrating the data in MSSQL, an object-relational database management system (RDBMS), which works at the backend, and the web interface was built in ASP.Net 2.0 with C#, HTML, Java Script & CSS.

The complete data are arranged below menu ProGlycProtdb under 10 broad fields that are further split into 47 subfields out of which 18 are content fields and 29 provide cross-references/links facilitating an easy access to existing information (Figure 2). The broad fields and subfields contain information for an entry as defined below:

ProGlycProt ID: a unique ProGlycProt ID that starts with series AC to indicate an archaeal characterized glycoprotein and BC to indicate bacterial characterized glycoprotein. Similarly, AU and BU series indicate archaeal and bacterial uncharacterized glycoproteins, respectively.

Organism information: contains general information about the source archaeal or bacterial species/strain.

Genome sequences: provides links to the available genome sequences and additional information like note on pathogenicity of source bacterial species/strain.

Gene information: enlists general information about the coding gene with relevant links.

Protein information: enlists name and other general information about the experimentally identified glycoprotein with relevant links.

Protein structure: provides available crystal structures or homology model with related links.

Glycosylation status: contains relevant links and information derived mainly from the literature about experimentally identified glycosites, type of glycosylation, experimental methods used to detect and define glycosites, a glycosite sequence logo and functional implication of glycosylation.

Glycan information: provides linear glycan structure (usually in standard IUPAC linear notation) corresponding BCSDB ID link and method of characterization of the glycan.

Protein glycosylation-linked gene(s): provides information about related, experimentally validated and predicted OSTs/GTs and relevant links

Literature: a tabulated bibliography and interesting additional information is given that could not be placed under aforesaid fields. For example, if a protein is glycoengineered or native, information about foreign OST

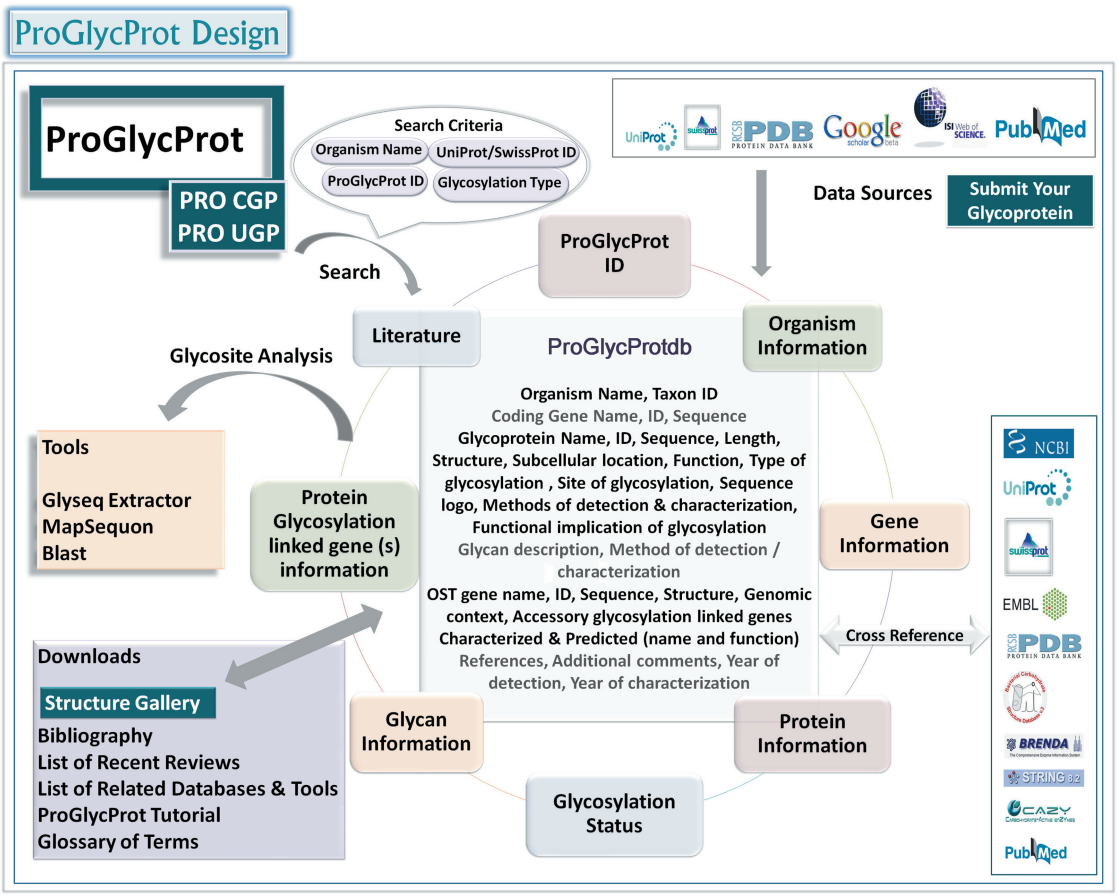

Figure 2. ProGlycProt data arrangement/retrieval schema.

used to glycosylate a protein of a given organism, sequon features, etc.

ProGlycProt is searchable by and for multiple parameters. A typical search result display (Supplementary Figure S3) and detailed note on data access is available as supplementary information.

TOOLS

A part of the literature in ProCGP, as discussed below, defines novel and potential sequon features in different bacterial glycoproteins belonging to different species. Some of these sequons are unique to prokaryotes. In the same context, there is a growing concern that existing glycosite prediction tools (as listed at http://www.proglycprot.org/related_tools_database.aspx) might not be sufficient or suitable for best analysis of prokaryotic glycoproteins (8). Interestingly, in a recent study by Comstock's group, in *Bacteroides fragilis* as many as eight new proteins have been characterized as glycoproteins, upon identification of the sequon (D)(S/T)(A/I/L/V/M/T) in corresponding sequences in bacterial proteome (23). The same group of researchers had validated this sequon experimentally while characterizing first *Bacteroides* glycoprotein BF2494 (24). Encouraged by this, we have developed tools Map Sequon (http://www.progpdb.org/Mapsequon.aspx) and Glyseq Extractor (http://www.progpdb.org/glyseq_extractor.aspx) that we believe can be of great help for making beginners' estimate of putative glycoproteins in prokaryotes, especially when one has to deal with proteome scale data. Map Sequon provides visual display and information about presence, spread or clustering of specified sequons in the input protein sequence(s). Similarly, Glyseq Extractor helps in retrieving defined sequence lengths around a sequon for statistical analysis of the glycosites. Based purely on the insights from the published literature irrespective of their statistical significance, the following sequons as found in native glycoproteins have been included in one or both the tools:

Typical in eukaryotes, **NX(S/T)** ($X \neq P$) sequon is required for N-glycosylation in glycoproteins of *Gammaproteobacteria* [HMW1 protein of *H. influenza*, (25)] as well as in almost all archaeal species (16). A recent characterization of PglB homolog of *Deltaproteobacteria Desulfovibrio desulfuricans* also suggests a preference for NX(S/T) sequon (26).

On the other hand, N-glycosylation at **(D/E)X_1NX(S/T)** (X_1 and $X \neq P$) sequon has almost always been found mediated by PglB protein (OST) of *Campylobacter* species and recently in case of *Helicobacter pullorum* that all belong to class *Epsilonproteobacteria* (27,28).

With currently available data, sequon **(D)(S/T)(A/I/L/V/M/T)** should be considered as an O-glycosylation feature exclusive to phylum *Bacteroidetes*. The sequon has an aspartate (D) preceding the glycosylated T or S which is followed by an amino acid with one or more methyl groups (24). The presence of this sequon has been observed consistently in glycoproteins of various members of this family belonging to all three but different classes, namely *Flavobacteria*, *Sphingobacteria* and *Bacteroidia*. One exception to this, Chondroitinase-B of *Pedobacter heparinus* lacks a methyl group containing amino acid at +1 position at the actual glycosylated sequon DSN (29) suggesting **DS** as a possible independent sequon feature that is supported in previous literature as well (30).

Similarly, glycosylation at tyrosine (Y) that is always preceded by valine (V) has been observed in all four sites of S-layer glycoprotein of *Thermoanaerobacter kivui* [original name *Acetogenium kivui*, phylum *Firmicutes* (31)], the first-available characterized glycoprotein with O-glycosylation at tyrosine. Therefore, we found it important to include **DS** as well as **VY** in our tool (s) to provide maximum coverage for possible sequons in prokaryotic glycoproteins.

The other common features observed around glycosites of O-glycosylated proteins of bacteria are S/T low complexity region at flexible-loop region of protein as in case of *N. gonorrhoeae* (32) and a eukaryotic mucin type Pro-, Ala-, Thr- and Ser-rich domains in *Actinobacteria* (33).

An additional tool BLAST (34) provides an easy retrieval of information using sequence similarity search against ProGlycProt. All these applications are accessible from ProGlycProt website under menu Tools.

WEB INTERFACE AND ADDITIONAL FEATURES

A free access to ProGlycProt database, tools and other features is available at http://www.proglycprot.org/. The curated data files, applications and additional features are arranged under four independent pull-down menus: ProGlycProtdb, Structure Gallery, Tools and Links. The browsing-enabled database statistics, our contact details and submission form for a new glycoprotein entry are available from the home page. A quick help is facilitated in the form of brief explanatory notes at the top of every page, explanatory text beneath various buttons, example display page and a detailed help section consisting of relevant FAQs, glossary of terms, and a downloadable tutorial on how to use ProGlycProt. Structure gallery renders independently an easy retrieval of crystal structures and homology models of characterized glycoproteins. Whereas a list of existing related databases/tools and a searchable bibliography and relevant recent reviews' list is available under links. An overall database design and flow of information in ProGlycProt is shown in Figure 2. More details on data access and print/download options under various menus are available as supplementary material.

CURRENT SCOPE AND FUTURE PERSPECTIVE

First release of ProGlycProt provides an extensive collection of experimentally identified prokaryotic glycosites (334), glycoproteins (95) and related information to set a stage for future statistical analysis of prokaryotic glycosites, neighbouring residues and 3D folds that can then provide fresh insights into the specificities of related OSTs and differences in the mechanisms of protein

glycosylation between prokaryotes and eukaryotes. For the reasons that ProGlycProt has a broad taxonomic coverage (Supplementary Figure S1) and published evidence of glycosylation for all entries, it provides an updated and realistic estimate of the extent of occurrence of protein glycosylation in prokaryotes. To serve a broader interest in prokaryotic glycoproteins, OSTs and associated GTs for their potential applied and basic applications (1–5,35), the database provides a variety of biologically and experimentally relevant information (Supplementary Table S1 and S2) about both native and glyco-engineered proteins of prokaryotes in addition to their cataloguing. Existing entries are updated in real time as soon as relevant literature is published or obtained. Otherwise, a general update policy is once in three months. The future versions aim at introducing in-depth information on prokaryotic OSTs along with continued compilation of characterized and uncharacterized glycoproteins under respective sections & enhanced structural/image inputs for glycan entries in ProCGP.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online: Supplementary Tables 1–3, Supplementary Figures 1–3.

ACKNOWLEDGEMENTS

Authors acknowledge Bulbul Khare for her contribution in data entries & statistics compilation.

FUNDING

Institute of Microbial Technology (OLP0063 to A.R.) and Council of Scientific and Industrial Research (SIP10AA to A.R.), India; the award of Junior Research Fellowship and Senior Research Fellowship by Council of Scientific and Industrial Research (to A.H.B. and J.S.C.); Institute of Microbial Technology for Research Internship grant (to H.M.). Funding for open access charge: Intramural research funds of Institute of Microbial Technology, Chandigarh.

Conflict of interest statement. None declared.

REFERENCES

1. Rosales-Borjas,D.M., Zambrano-Villa,S., Elinos,M., Kasem,H., Osuna,A., Mancilla,R. and Ortiz-Ortiz,L. (1998) Rapid screening test for tuberculosis using a 38-kDa antigen from *Mycobacterium tuberculosis*. *J. Clin. Lab. Anal.*, **12**, 126–129.
2. Roy,K., Hamilton,D., Ostmann,M.M. and Fleckenstein,J.M. (2009) Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic *Escherichia coli* in a murine model. *Vaccine*, **27**, 4601–4608.
3. Schäffer,C. and Messner,P. (2004) Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology*, **14**, 31R–42R.
4. Lloyd,R.C., Davis,B.G. and Jones,J.B. (2000) Site-selective glycosylation of subtilisin *Bacillus lentus* causes dramatic increases in esterase activity. *Bioorg. Med. Chem.*, **8**, 1537–1544.
5. Meldgaard,M. and Svendsen,I. (1994) Different effects of *N*-glycosylation on the thermostability of highly homologous bacterial (1,3-1,4)-beta-glucanases secreted from yeast. *Microbiology*, **140(Pt 1)**, 159–166.
6. Stepper,J., Shastri,S., Loo,T.S., Preston,J.C., Novak,P., Man,P., Moore,C.H., Havlicek,V., Patchett,M.L. and Norris,G.E. (2011) Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett.*, **585**, 645–650.
7. Nothaft,H. and Szymanski,C.M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.*, **8**, 765–778.
8. Dell,A., Galadari,A., Sastre,F. and Hitchen,P. (2010) Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int. J. Microbiol.*, **2010**, 148–178.
9. Upreti,R.K., Kumar,M. and Shankar,V. (2003) Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics*, **3**, 363–379.
10. Ragas,A., Roussel,L., Puzo,G. and Riviere,M. (2007) The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A. *J. Biol. Chem.*, **282**, 5133–5142.
11. Hirai,H., Takai,R., Iwano,M., Nakai,M., Kondo,M., Takayama,S., Isogai,A. and Che,F.S. (2011) Glycosylation regulates specific induction of rice immune responses by *Acidovorax avenae* Flagellin. *J. Biol. Chem.*, **286**, 25519–25530.
12. Jennings,M.P., Jen,F.E., Roddam,L.F., Apicella,M.A. and Edwards,J.L. (2011) *Neisseria gonorrhoeae* pilin glycan contributes to CR3 activation during challenge of primary cervical epithelial cells. *Cell. Microbiol.*, **13**, 885–896.
13. Moens,S. and Vanderleyden,J. (1997) Glycoproteins in prokaryotes. *Arch. Microbiol.*, **168**, 169–175.
14. Messner,P. and Schäffer,C. (2003) Prokaryotic glycoproteins. *Fortschr. Chem. Org. Naturst.*, **85**, 51–124.
15. Abu-Qarn,M., Eichler,J. and Sharon,N. (2008) Not just for Eukarya anymore: protein glycosylation in bacteria and archaea. *Curr. Opin. Struct. Biol.*, **18**, 544–550.
16. Calo,D., Kaminski,L. and Eichler,J. (2010) Protein glycosylation in archaea: sweet and extreme. *Glycobiology*, **20**, 1065–1076.
17. Gupta,R., Birch,H., Rapacki,K., Brunak,S. and Hansen,J.E. (1999) O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res.*, **27**, 370–372.
18. Jung,E., Veuthey,A.L., Gasteiger,E. and Bairoch,A. (2001) Annotation of glycoproteins in the SWISS-PROT database. *Proteomics*, **1**, 262–268.
19. Lee,T.Y., Huang,H.D., Hung,J.H., Huang,H.Y., Yang,Y.S. and Wang,T.H. (2006) dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res.*, **34**, D622–627.
20. Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Federhen,S. *et al.* (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **38**, D5–16.
21. Zeitler,R., Hochmuth,E., Deutzmann,R. and Sumper,M. (1998) Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent N-glycosylation of the cell surface glycoprotein from *Halobacterium halobium*. *Glycobiology*, **8**, 1157–1164.
22. Toukach,P.V. (2011) Bacterial carbohydrate structure database 3: principles and realization. *J. Chem. Inf. Model*, **51**, 159–170.
23. Fletcher,C.M., Coyne,M.J. and Comstock,L.E. (2011) Theoretical and experimental characterization of the scope of protein O-glycosylation in *Bacteroides fragilis*. *J. Biol. Chem.*, **286**, 3219–3226.
24. Fletcher,C.M., Coyne,M.J., Villa,O.F., Chatzidaki-Livanis,M. and Comstock,L.E. (2009) A general O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell*, **137**, 321–331.
25. Choi,K.J., Grass,S., Paek,S., St Geme,J.W. III and Yeo,H.J. (2010) The *Actinobacillus pleuropneumoniae* HMW1C-like glycosyltransferase mediates N-linked glycosylation of the *Haemophilus influenzae* HMW1 adhesin. *PLoS One*, **5**, e15888.
26. Ielmini,M.V. and Feldman,M.F. (2011) *Desulfovibrio desulfuricans* PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology*, **21**, 734–742.

27. Kowarik,M., Young,N.M., Numao,S., Schulz,B.L., Hug,I., Callewaert,N., Mills,D.C., Watson,D.C., Hernandez,M., Kelly,J.F. *et al*. (2006) Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J*., **25**, 1957–1966.
28. Jervis,A.J., Langdon,R., Hitchen,P., Lawson,A.J., Wood,A., Fothergill,J.L., Morris,H.R., Dell,A., Wren,B. and Linton,D. (2010) Characterization of N-linked protein glycosylation in *Helicobacter pullorum*. *J. Bacteriol*., **192**, 5228–5236.
29. Huang,W., Matte,A., Li,Y., Kim,Y.S., Linhardt,R.J., Su,H. and Cygler,M. (1999) Crystal structure of chondroitinase B from *Flavobacterium heparinum* and its complex with a disaccharide product at 1.7 A resolution. *J. Mol. Biol*., **294**, 1257–1269.
30. Plummer,T.H. Jr, Tarentino,A.L. and Hauer,C.R. (1995) Novel, specific O-glycosylation of secreted Flavobacterium meningosepticum proteins. Asp-Ser and Asp-Thr-Thr consensus sites. *J. Biol. Chem*., **270**, 13192–13196.
31. Peters,J., Rudolf,S., Oschkinat,H., Mengele,R., Sumper,M., Kellermann,J., Lottspeich,F. and Baumeister,W. (1992) Evidence for tyrosine-linked glycosaminoglycan in a bacterial surface protein. *Biol. Chem. Hoppe Seyler*, **373**, 171–176.
32. Vik,A., Aas,F.E., Anonsen,J.H., Bilsborough,S., Schneider,A., Egge-Jacobsen,W. and Koomey,M. (2009) Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci. USA*, **106**, 4447–4452.
33. Dobos,K.M., Khoo,K.H., Swiderek,K.M., Brennan,P.J. and Belisle,J.T. (1996) Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *J. Bacteriol*., **178**, 2498–2506.
34. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215**, 403–410.
35. Schwarz,F., Huang,W., Li,C., Schulz,B.L., Lizak,C., Palumbo,A., Numao,S., Neri,D., Aebi,M. and Wang,L.X. (2010) A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat. Chem. Biol*., **6**, 264–266.